



Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением

А. В. Сацюк^а, канд. техн. наук, доцент, orcid.org/0009-0006-7228-8279

Н. В. Володарец^б, канд. техн. наук, доцент, orcid.org/0000-0002-8526-4800, volodarets.nikita@yandex.ru

^аДонецкий институт железнодорожного транспорта, Горная ул., 6, Донецк, Донецкая Народная Республика, 283018, РФ

^бДонецкая академия транспорта, Дзержинского пр., 7, Донецк, Донецкая Народная Республика, 283086, РФ

Введение: современные системы компьютерного зрения для БПЛА сталкиваются с проблемой надежной детекции и трекинга в реальном времени при ограниченных ресурсах embedded-платформ, особенно при интеграции нейросетевых детекторов с алгоритмами сопровождения. Существующие реализации YOLO, несмотря на популярность, имеют недостатки: избыточную вычислительную сложность из-за ориентации на многоклассовую детекцию и неоптимальное взаимодействие с трекинг-алгоритмами. **Цель:** разработать оптимизированную версию YOLOv8 для гибридной системы детекции и трекинга на борту БПЛА с автоматическим наведением, направленную на снижение вычислительной сложности при сохранении точности и адаптацию к ресурсограниченным платформам. **Результаты:** исследование основано на экспериментах с модифицированной YOLOv8m, оцененной на embedded-платформе (Raspberry Pi 5) и специализированном датасете ONE_OBJECT. Разработана модифицированная версия YOLOv8m с избирательной заменой стандартных сверточных слоев на depthwise separable convolutions в блоках СЗСА и части слоев Neck. В результате экспериментальных исследований достигнуто снижение вычислительной сложности на 32,9 % (с 8,5 до 5,7 GFLOPS), количества параметров на 37,1 % (с 25,9 до 16,3 млн) и требований к памяти на 29,4 % (с 102 до 72 МБ). Скорость обработки на Raspberry Pi 5 увеличилась на 63,6 % (с 11 до 18 FPS) при сохранении высокой точности детекции mAP@0.5 на уровне 93,5 % — падение всего на 0,7 процентного пункта относительно базовой модели при тестировании на датасете ONE_OBJECT. Наибольшее снижение точности (1,2 процентного пункта) наблюдается для объектов малого размера (<50 пикселей). **Практическая значимость:** разработанная модификация YOLOv8m успешно интегрирована с алгоритмом трекинга CSRT. Следовательно, появляется возможность создавать эффективные гибридные системы автоматического наведения для БПЛА. Предложенное решение особенно перспективно для embedded-систем с ограниченными ресурсами. Более того, полученные результаты открывают новые возможности для создания энергоэффективных систем компьютерного зрения реального времени.

Ключевые слова — YOLOv8, depthwise separable convolutions, БПЛА, гибридная детекция, трекинг CSRT, модификация, оптимизация нейросетей.

Для цитирования: Сацюк А. В., Володарец Н. В. Модификация модели YOLO для гибридной системы детекции и трекинга в БПЛА с автоматическим наведением. *Информационно-управляющие системы*, 2025, № 4, с. 36–44. doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

For citation: Satsiuk A. V., Volodarets N. V. Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 36–44 (In Russian). doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

Введение

Сегодня активно развиваются интеллектуальные системы во многих отраслях. Так, на транспорте многочисленные исследования посвящены вопросам управления транспортными средствами, включая автоматическое регулирование движения транспортных средств [1, 2], разработку систем автономного управления [3–7], а также оптимизацию параметров, влияющих на движение [8–12]. Особого подхода в решении данных вопросов требуют системы управления беспилотными летательными аппаратами (БПЛА) [9].

Современные системы компьютерного зрения для БПЛА сталкиваются с комплексной задачей

обеспечения надежной детекции и сопровождения объектов в реальном времени при ограниченных вычислительных ресурсах [13–15]. Как показано в работе [16], алгоритм CSRT демонстрирует высокую эффективность в задачах трекинга, однако его работоспособность критически зависит от качества начальной детекции и периодической коррекции положения цели. Это выдвигает особые требования к нейросетевым моделям детекции, которые должны обеспечивать не только высокую точность распознавания, но и эффективную работу в составе гибридной системы.

Анализ современных исследований [1, 2, 17–20] показывает, что существующие реализации YOLO, несмотря на свою популярность, об-

ладают рядом ограничений при использовании в подобных системах. В работах [21, 22] отмечается проблема избыточной вычислительной сложности стандартных архитектур, что делает их неоптимальными для embedded-систем БПЛА. Исследование [23] выявило снижение точности детекции при значительном изменении масштаба объекта, что особенно критично при приближении БПЛА к цели. Авторы статьи [24] подчеркивают трудности интеграции детекторов с алгоритмами трекинга, связанные с неоптимальным форматом выходных данных и отсутствием механизмов адаптации к специфике работы трекера.

Между тем в работах [25, 26] рассмотрен механизм depthwise separable convolutions (DWS) как перспективный метод оптимизации вычислительной нагрузки в сверточных нейросетях. В исследовании [25] доказано, что замена стандартных сверток на DWS-аналоги в детекторах на основе YOLO сокращает параметры модели в 4–11 раз при сохранении 95 % точности. Авторы [26] отмечают, что DWS особенно эффективны для embedded-систем, обеспечивая до 30 % ускорения инференса. В [27] предложена адаптация DWS для задач монообъектной детекции, что актуально для систем с предопределенным целевым классом.

Таким образом, ключевой проблемой нейросетевых моделей детекции в гибридных системах БПЛА является неоптимальное использование вычислительных ресурсов. Как показывают исследования [22, 25], стандартные архитектуры содержат значительную избыточность параметров, обусловленную ориентацией на многоклассовую детекцию, тогда как в рассматриваемом случае требуется распознавание лишь единичного целевого объекта. Это приводит к необоснованному увеличению времени обработки и энергопотребления, что критично для embedded-систем с ограниченными ресурсами. При этом, как отмечается в [26, 27], существующие методы оптимизации, включая DWS, позволяют значительно сократить вычислительную сложность без ощутимой потери точности при работе с предопределенным классом объектов.

Целью авторов настоящей статьи является разработка модифицированной версии YOLOv8m, оптимизированной для работы в составе гибридной системы детекции и трекинга [16] на борту БПЛА с автоматическим наведением. Основные задачи включают сокращение вычислительной сложности модели при сохранении точности детекции, повышение устойчивости к изменению масштаба объекта, обеспечение совместимости с алгоритмом трекинга CSRT, а также адаптацию модели для работы на ресурсоограниченных embedded-платформах.

Выбор и обоснование базовой архитектуры. Математическая модель и алгоритмическая оптимизация архитектуры

В качестве базовой модели для модификации была выбрана YOLOv8m (medium-версия), что обусловлено оптимальным балансом между точностью детекции ($mAP@0.5 = 91\%$ на COCO) и вычислительной эффективностью (скорость обработки FPS $\sim 3,2$ мс на Tesla T4). Данная архитектура демонстрирует лучшие показатели FPS/точность среди семейства YOLOv8 при работе на ресурсоограниченных устройствах, что подтверждено бенчмарками в исследованиях [28–30]. Особое значение имеет модульная структура YOLOv8m, позволяющая производить точечные изменения backbone- и head-частей сети без критического нарушения целостности архитектуры.

В рамках данного исследования проведена корректировка архитектуры YOLOv8m для ее эффективной работы в составе гибридной системы детекции и трекинга на платформе Raspberry Pi 5. Основной акцент сделан на адаптацию сети под задачу детекции единичного объекта, что соответствует требованиям трекера CSRT, работающего только с одним целевым объектом, имеющим наивысший уровень достоверности. Такой подход позволил значительно снизить вычислительную нагрузку на систему без существенного ухудшения точности сопровождения.

Ключевым аспектом модернизации архитектуры YOLOv8m является замена части стандартных сверточных слоев на DWS (рис. 1, а и б). Данный метод, основанный на разделении операции свертки на два последовательных этапа: depthwise (пространственная свертка) и pointwise (объединение каналов), — продемонстрировал значительную эффективность в контексте ресурсоограниченных систем.

Математически процедура DWS выражается как композиция двух операций.

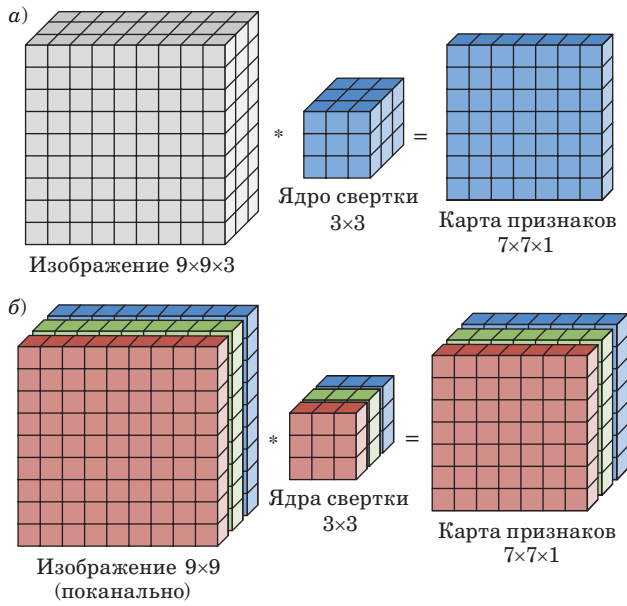
Первая — пространственная свертка, применяемая независимо к каждому каналу входного тензора:

$$DeptConv(W, y)_{(i,j)} = \sum_{k,l}^{K,L} w(k, l) \cdot y(i+k, j+l), \quad (1)$$

где w — двухмерное ядро свертки размерностью $K \times L$; y — входной тензор высотой и шириной $H \times W$ и с числом каналов C_{in} ; i, j — координаты выходного элемента; k, l — индексы перемещения по ядру w .

Вторая — свертка 1×1 , комбинирующая каналы:

$$PointConv(W, y)_{(i,j)} = \sum_m^{C_{in}} w_m \cdot y(i, j, m), \quad (2)$$



■ **Рис. 1.** Применения ядра свертки 3×3 с использованием классической сверточной обработки (а) и depthwise-обработки (б)

■ **Fig. 1.** Applications of 3×3 convolution kernel using classical convolutional processing (a) and depthwise processing (b)

где w_m — одномерное ядро свертки длиной M (число входных каналов) $\times N$ (число выходных каналов); i, j — координаты пространственного положения на карте признаков; m — индекс входного канала: $m \in [1, C_{in}]$; $y(i, j, m)$ — значение входного тензора в точке (i, j) канала m .

В стандартных сверточных слоях YOLO ядро фильтра размером $K \times K$ применяется ко всем C_{in} входным каналам. Тогда каждое ядро свертки имеет размер $K \times K \times C_{in}$ (на рис. 1, а — $9 \times 9 \times 3$), а выходной тензор имеет $(K - 1) \times (K - 2) \times C_{out}$ каналов [31].

В этом случае вычислительная сложность O_{conv} равна

$$O_{conv} = C_{in} \cdot C_{out} \cdot K^2 \cdot H \cdot W. \quad (3)$$

Таким образом, для входного изображения $C_{in} = 3$ (RGB-каналы), а в промежуточных слоях, например при входном тензоре размером $256 \times 56 \times 56$ ($C_{in} = 256$, $H = W = 56$), применение свертки с $C_{out} = 512$ фильтров размером 3×3 даст выходной тензор $512 \times 56 \times 56$. Тогда, согласно формуле (3), для данного случая $O_{conv} \approx 3,7 \cdot 10^9$.

В ситуации замены слоев со стандартной сверткой на DWS нет умножения на C_{out} , поскольку каждый канал обрабатывается независимо [25]. Поэтому формула сложности вычислений (3) примет вид

$$O_{DWS_S} = C_{in} \cdot K^2 \cdot H \cdot W \approx 7,2 \cdot 10^6. \quad (4)$$

С учетом применения точечной свертки (pointwise convolution), которая выполняет линейное смешивание каналов, каждый из 512 фильтров применяется ко всем 256 каналам в одной точке (H, W). В результате поэлементного умножения и последующего суммирования получается один новый канал [32]. При этом коэффициент $K^2 = 1^2 = 1$. Вычислительная сложность с учетом этого будет составлять

$$O_{DWS_P} = C_{in} \cdot C_{out} \cdot H \cdot W \approx 4,1 \cdot 10^8. \quad (5)$$

Для сравнения вычислительной сложности обычной свертки и DWS используем соотношение

$$\frac{O_{conv}}{O_{DWS_S} + O_{DWS_P}} = \frac{C_{out} \cdot K_{conv}^2}{K_{DWS_S}^2 + K_{DWS_P}^2} \cdot C_{out} \approx 8,8. \quad (6)$$

Из вычислений по формуле (6) видно, что DWS сокращает вычисления в 8,8 раза по сравнению с классической сверткой, разделяя операцию на depthwise-обработку каналов и pointwise-смешивание, что сохраняет точность при значительном ускорении.

Также замечено, что при использовании DWS с ядром 3×3 наблюдалось сокращение количества параметров в среднем в 1,73 раза (с 25,9 до 14,9 млн для типового слоя), что подтверждает перспективность данного подхода для embedded-решений.

Структурные изменения модифицированной YOLOv8m

При модификации YOLOv8m была применена избирательная стратегия замены слоев: преобразованию подверглись преимущественно слои, обрабатывающие признаки среднего и высокого уровня абстракции, где отмечалась наибольшая избыточность вычислительных операций. Важно отметить, что начальные сверточные слои, ответственные за первичную обработку изображения и формирование базовых признаков, были сохранены в исходном виде. Такой подход позволил сохранить информативность низкоуровневых признаков, критически важных для последующих этапов детекции и трекинга.

Выбор конкретных слоев для замены на DWS основывался на результатах комплексного экспериментального исследования. На первом этапе проводился анализ распределения вычислительной нагрузки по слоям модели с использованием

инструментария TorchProfiler. Были идентифицированы слои, вносящие максимальный вклад в общее количество операций (свыше 15 % от 8,5 GFLOPS исходной модели). Затем для этих слоев выполнялось пошаговое ablation-исследование: последовательная замена каждого кандидатного слоя на DWS-аналог с оценкой влияния на mAP@0.5 и FPS на тестовой выборке датасета ONE_OBJECT.

Как показали результаты, замена слоев в блоках C3CA_3 и C3CA_9 позволяла достичь 28–35 % сокращения вычислительной сложности при потере точности не более 0,3 процентного пункта для каждого модифицированного слоя. При этом модификация начальных сверточных слоев приводила к существенному падению mAP@0.5 (до 2,5 %), что подтвердило их критическую важность для формирования базовых признаков. Дополнительным критерием отбора служила устойчивость к изменению масштаба объекта — слои, демонстрирующие наименьшее ухудшение точности для объектов малого размера (<50 пикселей), получали приоритет при замене.

Реализация модификаций осуществлялась средствами PyTorch с сохранением совместимости с оригинальной архитектурой YOLOv8 [33]. Для заменяемых слоев применялась специальная процедура инициализации весов: параметры depthwise-части формировались путем декомпозиции предобученных ядер исходной модели,

в то время как pointwise-компоненты инициализировались случайным образом с последующим обучением на целевом датасете ONE_OBJECT (820 изображений целевого объекта, один класс).

С учетом вышеописанного модифицированная архитектура YOLOv8m, где ключевым изменением стала замена стандартных сверточных слоев на DWS в определенных блоках, представлена на рис. 2.

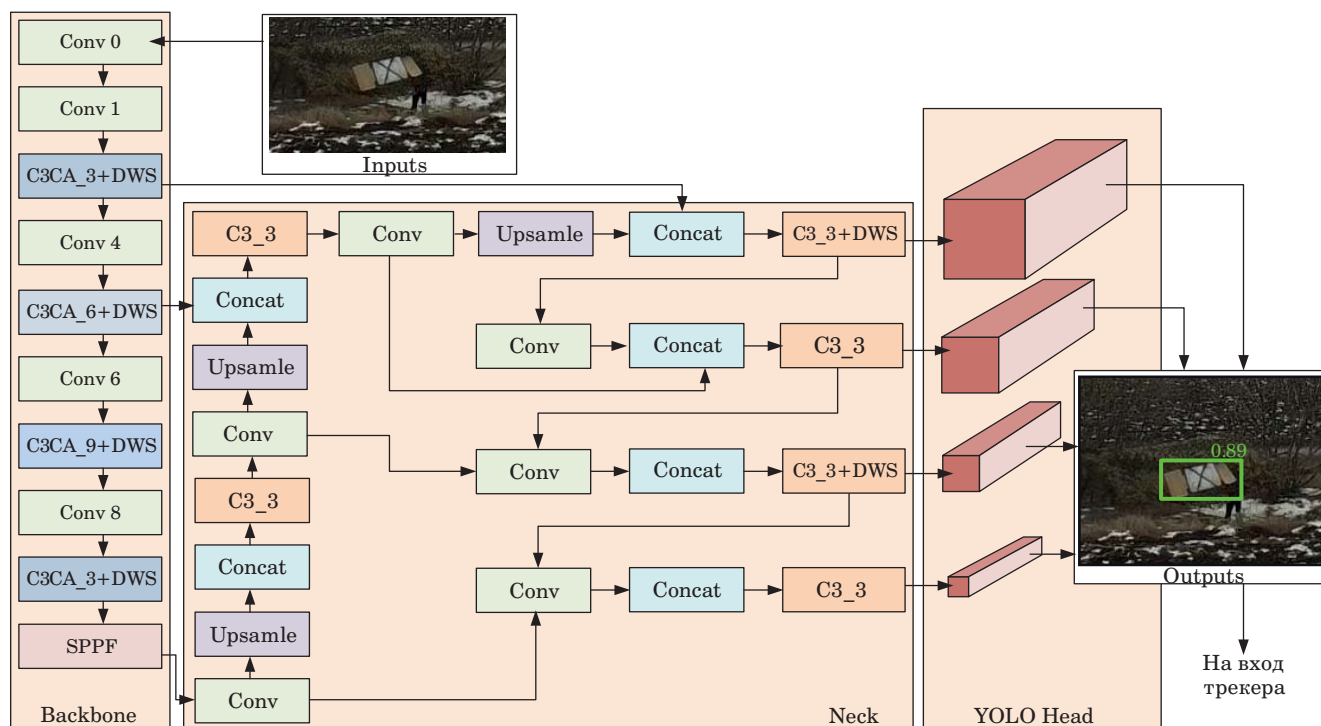
Особенностью предложенной структуры является использование специализированных блоков C3CA [34] в сочетании с DWS-свертками.

Основные преобразования затронули следующие блоки:

- блоки C3CA_3 подверглись частичной замене (50 %) внутренних сверток на DWS-аналоги;
- блоки C3CA_6 получили полную замену стандартных сверток на DWS-версии;
- блоки C3CA_9 получили полную замену стандартных сверток на DWS-версии;
- промежуточные сверточные слои (Conv 1, Conv 4, Conv 6, Conv 8) оптимизированы с сохранением базовой функциональности.

Входной сверточный слой (блок Conv 0) оставлен без изменений для сохранения низкоуровневых признаков. Финальный SSPF-блок остался без изменений для сохранения пространственной пирамидальной структуры.

В ходе оптимизации архитектуры модификации подвергся также блок Neck, отвечающий



■ **Рис. 2.** Структура модифицированной YOLOv8m

■ **Fig. 2.** Modified structure of YOLOv8m

за агрегацию и перераспределение признаков различного масштаба. Основные изменения включают избирательную замену стандартных сверточных слоев на DWS с сохранением критически важных элементов структуры. В верхних слоях Neck, получающих высокоуровневые признаки из Backbone, оставлены обычные свертки для сохранения информативности признаков. Средние и нижние слои, где требования к точности менее строги, были оптимизированы через внедрение DWS-аналогов, что позволило сократить количество параметров на 40 % для каждого модифицированного слоя. Особое внимание уделено блокам C3_3, где DWS-свертки применены только в residual-ветвях при сохранении основной shortcut-ветви неизменной для стабильности обучения. Первый и последний сверточные слои Neck не подвергались модификации, поскольку они играют ключевую роль в инициализации потока признаков и формировании выходных данных для YOLO Head.

Результаты исследования

Для валидации модифицированной архитектуры использовался специализированный датасет ONE_OBJECT, содержащий 820 изображений целевого объекта в разрешении 1920×1080 пикселей. Датасет был создан на основе аэрофото-съемки с БПЛА, выполненной с различных ракурсов и расстояний от 5 до 400 м. В процессе съемки специально варьировалось фоновое окружение для повышения вариативности данных. Особое внимание при формировании датасета уделялось включению объектов малого размера (<50 пикселей), что составило 18 % от общего количества образцов и позволило оценить масштабную устойчивость модели.

Первичные видеоматериалы подвергались комплексной постобработке, включавшей геометрические преобразования с поворотами до $\pm 15^\circ$ и масштабированием $\pm 20\%$, цветокоррек-

цию с вариацией яркости $\pm 30\%$ и контрастности $\pm 25\%$, а также добавление гауссова шума с $\sigma = 0,05$. Такой подход к аугментации данных обеспечил необходимое разнообразие обучающей выборки.

Экспериментальные исследования модифицированной версии YOLOv8m проводились с оценкой по ключевым показателям (таблица).

Результаты экспериментального исследования демонстрируют эффективность предложенных архитектурных изменений при сохранении приемлемого уровня точности детекции (рис. 3).

Сравнительный анализ стандартной и модифицированной версий YOLOv8m на датасете ONE_OBJECT выявил следующие закономерности. По метрике mAP@0.5 наблюдалось незначительное снижение точности на 0,7 процентного пункта (с 94,2 до 93,5 %), что находится в пределах статистической погрешности для задач моно-объектной детекции. Наибольшее расхождение в 1,2 % зафиксировано для случаев детекции объектов малого размера (<50 пикселей), что объясняется уменьшением рецептивного поля в оптимизированных слоях.

По производительности модифицированная архитектура показала существенное улучшение ключевых показателей. Скорость обработки возросла на 63,6 % (с 11 до 18 FPS) при тестировании на платформе Raspberry Pi5, что напрямую связано со снижением вычислительной сложности на 32,9 % (с 8,5 до 5,7 GFLOPS).

Анализ ресурсоемкости выявил сокращение количества параметров модели на 37,1 % (с 25,9 до 16,3 млн) и уменьшение требований к памяти на 29,4 % (со 102 до 72 МБ). Эти изменения особенно значимы для систем с ограниченными вычислительными возможностями, где каждый мегабайт памяти имеет критическое значение.

Результаты тестирования также показали снижение среднего энергопотребления системы с 4,8 до 4,4 Вт (–8,3 %) при работе на Raspberry Pi5. Это незначительное улучшение энергоэффективности достигнуто благодаря оптимиза-

- Сравнения моделей YOLOv8m
- Comparison of YOLOv8m models

Параметр	Стандартная YOLOv8m	Модифицированная YOLOv8m	Изменение, %
mAP@0.5	94,2 %	93,5 %	–0,7
FPS (Raspberry Pi5)	11	18	+63,6
GFLOPS	8,5	5,7	–32,9
Количество параметров модели	25,9 млн	16,3 млн	–37,1
Память	102 МБ	72 МБ	–29,4
Средняя энергия потребления	4,8 Вт	4,4 Вт	–8,3



■ **Рис. 3.** Результат сравнительного эксперимента стандартной (слева) и модифицированной (справа) YOLOv8m

■ **Fig. 3.** The result of a comparative experiment of standard (left) and modified (right) YOLOv8m

ции вычислительной нагрузки модели, что особенно важно для автономных систем с ограниченными энергетическими ресурсами, таких как бортовые компьютеры БПЛА.

Литература

1. Hanzla M., Jalal A. Intelligent transportation surveillance via YOLOv9 and NASNet over aerial imagery. 2025 6th International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, 2025, pp. 1–8. doi:10.1109/ICACS64902.2025.10937840
2. Chen J., Wang Q., Cheng H. H., Peng W., Xu W. A review of vision-based traffic semantic understanding in ITSs. *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 23, iss. 11, pp. 19954–19979. doi:10.1109/TITS.2022.3182410
3. Ganesan M., Chokkalingam B., Kandhasamy S. Implementation of different road user detection with custom dataset using deep learning algorithm for autonomous vehicle. 2024 IEEE 4th International Conference on Sustainable Energy and Future Electric

Заключение

Проведенная модификация YOLOv8m путем замены стандартных сверточных слоев на DWS и оптимизации блоков СЗСА позволила создать эффективную модель для детекции единичного объекта в системах реального времени для БПЛА. Эксперименты показали, что предложенная архитектура сокращает вычислительную нагрузку на 32,9 % (5,7 GFLOPS) и уменьшает объем параметров на 37,1 % (16,3 млн) при сохранении точности детекции (mAP@0.5 = 93,5 %). Наибольшее снижение точности на 1,2 процентного пункта наблюдается для объектов малого размера (<50 пикселей), что открывает перспективы для дальнейшей оптимизации в этом направлении. Также представляет интерес исследование возможностей интеграции с другими алгоритмами трекинга и адаптации архитектуры для задач мультиобъектного сопровождения. Полученные результаты демонстрируют, что избирательное применение DWS-сверток в слоях среднего уровня обеспечивает оптимальный баланс между производительностью и качеством работы модели. Модифицированная версия YOLOv8m особенно перспективна для embedded-систем с ограниченными ресурсами, где критически важны энергоэффективность и скорость обработки, а также открывает новые возможности для создания легковесных гибридных систем детекции и трекинга в различных приложениях компьютерного зрения.

Финансовая поддержка

Исследование выполнено за счет гранта Российского научного фонда № 25-21-20601, <https://rscf.ru/project/25-21-20601/>.

Transportation (SEFET), Hyderabad, India, 2024, pp. 1–7. doi:10.1109/SEFET61574.2024.10718012

4. Rajabli N., Flammini F., Nardone R., Vittorini V. Software verification and validation of safe autonomous cars: A systematic literature review. *IEEE Access*, 2021, vol. 9, pp. 4797–4819. <https://doi.org/10.1109/ACCESS.2020.3048047>
5. Wang K., Shen C., Li X., Lu J. Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2025, vol. 26, no. 3, pp. 2880–2896. doi:10.1109/TITS.2025.3532803
6. Иванов М. Е., Матасов Ю. Ф. Применение нейросетевого регулятора для повышения качества переходного процесса системы стабилизации высоты полета экраноплана. *Информационно-управляющие системы*, 2024, № 6, с. 27–34. doi:10.31799/1684-8853-2024-6-27-34, EDN: BOBWBW

7. Luu T. A. Application of unmanned aerial vehicle in topographical surveying: A review. *Sustainable Development of Mountain Territories*, 2024, vol. 16, no. 4, pp. 1833–1848. doi:10.21177/1998-4502-2024-16-4-1833-1848
8. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Volkov V., Ahieiev M., Pohorletskyi D., Zinchenko S., Khudiakov I. Development of a simulation model of a system for automatic regulation of the capacity of a compressor unit. *SAE Technical Papers*, Warrendale, Warrendale, PA, 2022, no. 2022-01-5022, p. 5022. doi:10.4271/2022-01-5022
9. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskyi D., Tkachuk K., Nosov P., Ukrainskyi Ye. Improvement of production processes of the marshalling yard in terms of automatic power control of the compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5063, p. 5063. doi:10.4271/2022-01-5063
10. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskyi D., Tkachuk K., Nosov P., Sumtsov A. Development of a multifactor model for determining energy costs at a compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5079, p. 5079. doi:10.4271/2022-01-5079
11. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podnebenna S., Belousov E., Ahieiev M., Pohorletskyi D., Serhii Z., Khudiakov I. Development of the structure for the technical implementation of the modernized system for automatic regulation of the capacity of the compressor unit at the marshalling yard. *SAE Technical Papers*, 2022, no. 2022-01-5084, p. 5084. doi:10.4271/2022-01-5084
12. Volodarets M., Gritsuk I., Ukrainskyi Y., Shein V., Stepanov O., Khudiakov I., Ahieiev M., Vychuzhanin V., Smyrnov O., Saraiev O. Development of the analytical system for vehicle operating conditions management in the V2I information complex using simulation modeling. *Eastern-European Journal of Enterprise Technologies*, 2020, no. 3 (107), pp. 6–16. doi:10.15587/1729-4061.2020.215006
13. Redmon J., Farhadi A. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. doi:10.48550/arXiv.1804.02767
14. Bochkovskiy A., Wang C.-Y., Liao H.-Y. M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. doi:10.48550/arXiv.2004.10934
15. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. doi:10.48550/arXiv.2207.02696
16. Сацюк А. В., Володарец Н. В. Оптимизация параметров трекера CSRT в системе слежения беспилотных летательных аппаратов. *Устойчивое развитие горных территорий*, 2025, т. 17, № 2. <https://naukagor.ru/ru-ru/about/textsofarticles> (дата обращения: 07 июля 2025). (В печати.)
17. Корилов А. М., Тран В. Т. Нейросетевой анализ видеопотока, поступающего с камеры беспилотного летательного аппарата, для компенсации ошибок системы навигации и управления БПЛА. *Телекоммуникации*, 2023, № 7, с. 13–23. doi:10.31044/1684-2588-2023-0-7-13-23
18. Viswanatha V., Chandana R. K., Ramachandra A. C. Real time object detection system with YOLO and CNN models: A review. *arXiv preprint arXiv:2208.773*, 2022. doi:10.48550/arXiv.2208.00773
19. Yuan K., Zhuang X., Schaefer G., Feng J., Guan L., Fang H. Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2021, vol. 14, pp. 7422–7434. doi:10.1109/JSTARS.2021.3098678
20. Gallagher J. E., Oughton E. J. Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications, and challenges. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), IEEE Access*, 2025, vol. 13, pp. 7366–7395. doi:10.1109/ACCESS.2025.3526458
21. Xu S., Xue Y. Lightweight object detection for UAVs using compressed feature fusion and attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, pp. 1–14. doi:10.1109/TGRS.2023.3267421
22. Satsuk A. V., Volodarets N. V., Krupskiy D. I. Acceleration of neural network for UAV operation in real time. *Вестник Донецкой академии транспорта*, 2025, № 1, с. 6–14.
23. Laghari A. A., Jumani A. K., Laghari R. A., Li H., Karim S., Khan A. A. Unmanned aerial vehicles advances in object detection and communication security review. *Cognitive Robotics*, 2024, vol. 4, pp. 73–82. doi:10.1016/j.cogr.2024.07.002
24. Wang Q., Gao J., Xing J., Zhang M. DETR4D: Transformer-based multi-object tracking for autonomous driving with depth-aware detection. *IEEE Robotics and Automation Letters*, 2023, vol. 8, no. 4, pp. 1931–1938. doi:10.1109/LRA.2023.3246839
25. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. doi:10.48550/arXiv.1704.04861
26. Hasan M. A., Dey K. Depthwise separable convolutions with deep residual convolutions. *arXiv preprint arXiv:2411.07544*, 2024. doi:10.48550/arXiv.2411.07544
27. Zhao X., Wang L., Zhang Y., Han X., Deveci M., Parmar M. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 2024, vol. 57, art. no. 99, p. 43. doi:10.1007/s10462-024-10721-6
28. Wang J.-W., Cao Y., Guo Z.-K., Xu C. Research on long-distance snow depth measurement method based on improved YOLOv8. *IEEE Access*, 2025, vol. 13, pp. 55370–55380. doi:10.1109/ACCESS.2025.3551727

29. Varghese R., Sambath M. Ultralytics YOLOv8: A novel object detection algorithm with enhanced performance and robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 1–6. doi:10.1109/ADICS58448.2024.10533619
30. Alqahtani D. K., Cheema M. A., Toosi A. N. Benchmarking deep learning models for object detection on edge computing devices. *International Conference on Service-Oriented Computing*, LNCS, Singapore, Springer Nature Singapore, 2025, vol. 15404, pp. 142–150. doi:10.1007/978-981-96-0805-8_11
31. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, pp. 436–444. doi:10.1038/nature14539
32. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
33. Wang X., Li Z., Yang J. PyTorch-YOLO-Nano. A framework for efficient deployment of modified YOLO architectures on embedded GPUs. *Journal of Real-Time Image Processing*, 2025, no. 22(1), pp. 45–60. doi:10.1007/s11554-024-01458-0
34. Chen L., Wang Z., Liu F. C3Fusion: Lightweight cross-channel and spatial attention for real-time object detection on edge devices. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, no. 46(5), pp. 1123–1136. doi:10.1109/TPAMI.2024.3385210

UDC 004.8+004.93+629.7

doi:10.31799/1684-8853-2025-4-36-44

EDN: YKQVJU

Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system

A. V. Satsiuk^a, PhD, Tech., Associate Professor, orcid.org/0009-0006-7228-8279

N. V. Volodarets^b, PhD, Tech., Associate Professor, orcid.org/0000-0002-8526-4800, volodarets.nikita@yandex.ru

^aDonetsk Institute of Railway Transport, 6, Gornaya St., 283018, Donetsk, Russian Federation

^bDonetsk Academy of Transport, 7, Dzerzhinsky Ave., 283086, Donetsk, Russian Federation

Introduction: Modern computer vision systems for UAVs face the problem of reliable detection and tracking in real time with limited resources of embedded platforms, especially when integrating neural network detectors with tracking algorithms. Existing YOLO implementations, despite their popularity, have drawbacks: excessive computational complexity due to the focus on multi-class detection and non-optimal interaction with tracking algorithms. **Purpose:** To develop an optimized version of YOLOv8 for a hybrid detection and an on-board tracking system for a UAV with automatic guidance, aimed at reducing computational complexity while maintaining accuracy and adapting to resource-limited platforms. **Results:** The study is based on experiments with a modified YOLOv8m, evaluated on an embedded platform (Raspberry Pi 5) and a specialized ONE_OBJECT dataset. We develop a modified version of YOLOv8m with selective replacement of standard convolutional layers with depthwise separable convolutions in C3CA blocks and some Neck layers. As a result of experimental studies, we achieve a reduction in computational complexity by 32.9% (from 8.5 to 5.7 GFLOPS), the number of parameters by 37.1% (from 25.9 million to 16.3 million) and memory requirements by 29.4% (from 102 to 72 MB). The processing speed on Raspberry Pi 5 increases by 63.6% (from 11 to 18 FPS) while maintaining high detection accuracy mAP@0.5 at 93.5% (a drop of only 0.7 percentage points relative to the base model when tested on the ONE_OBJECT dataset). The greatest decrease in accuracy (1.2 percentage points) is observed for small objects (less than 50 pixels). **Practical relevance:** The developed YOLOv8m modification has been successfully integrated with the CSRT tracking algorithm. Consequently, it becomes possible to create efficient hybrid automatic guidance systems for UAVs. The proposed solution is especially promising for embedded systems with limited resources. In addition, the obtained results open up new possibilities for creating energy-efficient real-time computer vision systems.

Keywords — YOLOv8, depthwise separable convolutions, UAV, hybrid detection, CSRT tracking, modification, neural network optimization.

For citation: Satsiuk A. V., Volodarets N. V. Modification of the YOLO model for a hybrid detection and tracking system in UAVs with an automatic guidance system. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2025, no. 4, pp. 36–44 (In Russian). doi:10.31799/1684-8853-2025-4-36-44, EDN: YKQVJU

Financial support

The study was supported by the Russian Science Foundation, grant No. 25-21-20601, <https://rscf.ru/project/25-21-20601/>.

Reference

1. Hanzla M., Jalal A. Intelligent transportation surveillance via YOLOv9 and NASNet over aerial imagery. *2025 6th International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, 2025, pp. 1–8. doi:10.1109/ICACS64902.2025.10937840
2. Chen J., Wang Q., Cheng H. H., Peng W., Xu W. A review of vision-based traffic semantic understanding in ITSs. *IEEE Transactions on Intelligent Transportation Systems*, 2022, vol. 23, iss. 11, pp. 19954–19979. doi:10.1109/TITS.2022.3182410
3. Ganesan M., Chokkalingam B., Kandhasamy S. Implementation of different road user detection with custom dataset using deep learning algorithm for autonomous vehicle. *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, Hyderabad, India, 2024, pp. 1–7. doi:10.1109/SEFET61574.2024.10718012
4. Rajabli N., Flammini F., Nardone R., Vittorini V. Software verification and validation of safe autonomous cars: A systematic literature review. *IEEE Access*, 2021, vol. 9, pp. 4797–4819. <https://doi.org/10.1109/ACCESS.2020.3048047>

5. Wang K., Shen C., Li X., Lu J. Uncertainty quantification for safe and reliable autonomous vehicles: A review of methods and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2025, vol. 26, no. 3, pp. 2880–2896. doi:10.1109/TITS.2025.3532803
6. Ivanov M. E., Matasov Yu. F. Neuro-controller for improving the quality of the transient process of the WIG-craft altitude stabilization system. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2024, no. 6, pp. 27–34 (In Russian). doi:10.31799/1684-8853-2024-6-27-34, EDN: BOBWN
7. Luu T. A. Application of unmanned aerial vehicle in topographical surveying: A review. *Sustainable Development of Mountain Territories*, 2024, vol. 16, no. 4, pp. 1833–1848. doi:10.21177/1998-4502-2024-16-4-1833-1848
8. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podneben-na S., Belousov E., Volkov V., Ahieiev M., Pohorletskyi D., Zinchenko S., Khudiakov I. Development of a simulation model of a system for automatic regulation of the capacity of a compressor unit. *SAE Technical Papers*, Warrendale, Warrendale, PA, 2022, no. 2022-01-5022, p. 5022. doi:10.4271/2022-01-5022
9. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskyi D., Tkachuk K., Nosov P., Ukrain-skyi Ye. Improvement of production processes of the mar-shalling yard in terms of automatic power control of the compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5063, p. 5063. doi:10.4271/2022-01-5063
10. Volodarets M., Satsiuk A., Gritsuk I., Litikova O., Ahieiev M., Belousov E., Pohorletskyi D., Tkachuk K., Nosov P., Sumstov A. Development of a multifactor model for determining energy costs at a compressor station. *SAE Technical Papers*, 2022, no. 2022-01-5079, p. 5079. doi:10.4271/2022-01-5079
11. Satsiuk A., Volodarets M., Gritsuk I., Litikova H., Podneben-na S., Belousov E., Ahieiev M., Pohorletskyi D., Serhii Z., Khudiakov I. Development of the structure for the technical implementation of the modernized system for automatic reg-ulation of the capacity of the compressor unit at the mar-shalling yard. *SAE Technical Papers*, 2022, no. 2022-01-5084, p. 5084. doi:10.4271/2022-01-5084
12. Volodarets M., Gritsuk I., Ukrainskyi Y., Shein V., Stepanov O., Khudiakov I., Ahieiev M., Vychuzhanin V., Smyrnov O., Saraiev O. Development of the analytical system for vehicle operating conditions management in the V2I information complex using simulation modeling. *Eastern-European Journal of Enterprise Technologies*, 2020, no. 3 (107), pp. 6–16. doi:10.15587/1729-4061.2020.215006
13. Redmon J., Farhadi A. YOLOv3: An incremental improve-ment. *arXiv preprint arXiv:1804.02767*, 2018. doi:10.48550/arXiv.1804.02767
14. Bochkovskiy A., Wang C.-Y., Liao H.-Y. M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. doi:10.48550/arXiv.2004.10934
15. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022. doi:10.48550/arXiv.2207.02696
16. Satsiuk A. V., Volodarets N. V. Optimization of CSRT track-er parameters in unmanned aerial vehicle tracking system. *Sustainable Development of Mountain Territories*, 2025, vol. 17, no. 2. Available at: <https://naukagor.ru/ru-ru/about/textsofarticles> (accessed 07 July 2025) (In Russian). (In print.)
17. Korikov A. M., Tran V. T. Neural network analysis of the video stream coming from the camera of an unmanned aerial vehicle to compensate for errors in the UAV naviga-tion and control system. *Telecommunications*, 2023, no. 7, pp. 13–23. doi:10.31044/1684-2588-2023-0-7-13-23
18. Viswanatha V., Chandana R. K., Ramachandra A. C. Real time object detection system with YOLO and CNN models: A review. *arXiv preprint arXiv:2208.773*, 2022. doi:10.48550/arXiv.2208.00773
19. Yuan K., Zhuang X., Schaefer G., Feng J., Guan L., Fang H. Deep-learning-based multispectral satellite image segmen-tation for water body detection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2021, vol. 14, pp. 7422–7434. doi:10.1109/JSTARS.2021.3098678
20. Gallagher J. E., Oughton E. J. Surveying You Only Look Once (YOLO) multispectral object detection advancements, applications, and challenges. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), IEEE Access*, 2025, vol. 13, pp. 7366–7395. doi:10.1109/ACCESS.2025.3526458
21. Xu S., Xue Y. Lightweight object detection for UAVs using compressed feature fusion and attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, pp. 1–14. doi:10.1109/TGRS.2023.3267421
22. Satsuk A. V., Volodarets N. V., Krupskiy D. I. Acceleration of neural network for UAV operation in real time. *Vestnik Do-neckoj Akademii Transporta*, 2025, no. 1, pp. 6–14.
23. Laghari A. A., Jumani A. K., Laghari R. A., Li H., Karim S., Khan A. A. Unmanned aerial vehicles advances in object de-tection and communication security review. *Cognitive Ro-botics*, 2024, vol. 4, pp. 73–82. doi:10.1016/j.cogr.2024.07.002
24. Wang Q., Gao J., Xing J., Zhang M. DETR4D: Transform-er-based multi-object tracking for autonomous driving with depth-aware detection. *IEEE Robotics and Automation Let-ters*, 2023, vol. 8, no. 4, pp. 1931–1938. doi:10.1109/LRA.2023.3246839
25. Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: Efficient convolutional neural networks for mobile vision applica-tions. *arXiv preprint arXiv:1704.04861*, 2017. doi:10.48550/arXiv.1704.04861
26. Hasan M. A., Dey K. Depthwise separable convolutions with deep residual convolutions. *arXiv preprint arXiv:2411.07544*, 2024. doi:10.48550/arXiv.2411.07544
27. Zhao X., Wang L., Zhang Y., Han X., Deveci M., Parmar M. A review of convolutional neural networks in computer vi-sion. *Artificial Intelligence Review*, 2024, vol. 57, art. no. 99, p. 43. doi:10.1007/s10462-024-10721-6
28. Wang J.-W., Cao Y., Guo Z.-K., Xu C. Research on long-dis-tance snow depth measurement method based on improved YOLOv8. *IEEE Access*, 2025, vol. 13, pp. 55370–55380. doi:10.1109/ACCESS.2025.3551727
29. Varghese R., Sambath M. Ultralytics Yolov8: A novel object detection algorithm with enhanced performance and ro-bustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Chennai, India, 2024, pp. 1–6. doi:10.1109/ADICS58448.2024.10533619
30. Alqahtani D. K., Cheema M. A., Toosi A. N. Benchmarking deep learning models for object detection on edge computing devices. *International Conference on Service-Oriented Com-puting*, LNCS, Singapore, Springer Nature Singapore, 2025, vol. 15404, pp. 142–150. doi:10.1007/978-981-96-0805-8_11
31. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, pp. 436–444. doi:10.1038/nature14539
32. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE Conference on Com-puter Vision and Pattern Recognition*, 2017, pp. 1251–1258.
33. Wang X., Li Z., Yang J. PyTorch-YOLO-Nano. A framework for efficient deployment of modified YOLO architectures on embedded GPUs. *Journal of Real-Time Image Processing*, 2025, no. 22(1), pp. 45–60. doi:10.1007/s11554-024-01458-0
34. Chen L., Wang Z., Liu F. C3Fusion: Lightweight cross-chan-nel and spatial attention for real-time object detection on edge devices. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023, no. 46(5), pp. 1123–1136. doi:10.1109/TPAMI.2024.3385210